Contribution of the tyrosines to the structure and function of the human U1A N-terminal RNA binding domain

JAMES K. KRANZ, JIRONG LU, AND KATHLEEN B. HALL

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine,
St. Louis, Missouri 63110

(RECEIVED April 2, 1996; ACCEPTED May 14, 1996)

Abstract

RNA binding domains (RBDs) are members of a large family of proteins that share minimal sequence conservation but adopt an $\alpha\beta$ sandwich global fold. Defining the contributions of specific amino acids to RBD structure and RNA binding is critical to understanding the functions of these proteins. In these experiments with the human U1A N-terminal RNA binding domain (RBD1), the contributions from each of its four tyrosines to protein structure, stability, and RNA binding were measured. Each tyrosine was substituted with phenylalanine and one other selected residue, and the resulting proteins were characterized by chemical denaturation to measure their unfolding free energy, by binding free energies to the wild-type RNA hairpin, and by ^{19}F NMR to probe for structural changes. Features of the protein identified in these experiments include a possible tyrosine/lysine contact in an α -helix, which may be an example of an energetically favorable aromatic/amino side chain interaction. One long loop of the protein, which shows unusual ^{15}N backbone and tyrosine side-chain dynamics, is implicated in protein:protein association. The diverse interactions of the four tyrosine residues in the organization of RBD1 illustrate how each member of this family of proteins will have unique molecular details that contribute to function.

Keywords: NMR; protein structure and energetics; RNA binding domain

The properties of the RNA binding domain, RNA recognition motif, or ribonucleoprotein (Bandziulis et al., 1989) are of interest first because these protein domains can specifically recognize a wide variety of RNA targets and second because they adopt an $\alpha\beta$ sandwich global topology that is one of the super-families of protein folds (Orengo & Thornton, 1993). As RBDs, they contain two conserved sequences that identify them as members of this family of RNA binding proteins (Dreyfuss et al., 1988; Bandziulis et al., 1989; Mattaj, 1989; Query et al., 1989; Kenan et al., 1991; Birney et al., 1993; see Keene & Query, 1991 and Dreyfuss et al., 1993 for reviews), as well as specific conserved residues that participate in formation of the $\beta\alpha\beta$ - $\beta\alpha\beta$ fold. As an $\alpha\beta$ sandwich, the alignment of each structural element to form the folded protein is critical to packing and stability. Through study of the contribution of specific amino acid

residues to RBD structure, stability, and RNA binding, it may be possible to correlate structure with function in this large family of proteins.

Several RBD structures have been described (Nagai et al., 1990; Hoffman et al., 1991; Wittekind et al., 1992; Garrett et al., 1994; Lee et al., 1994; Lu & Hall, 1995), and all share a common folding topology. The $\beta\alpha\beta$ - $\beta\alpha\beta$ secondary structure forms a four-stranded antiparallel β -sheet, the two α -helices lie on one side, and the interface formed between the helices and β -sheet constitutes the hydrophobic core of the protein. This structure is illustrated in Figure 1 for the N-terminal RBD1 of the human U1A protein. The β strands are arranged in the order $\beta 4\beta 1\beta 3\beta 2$, with the conserved $\beta 1$ (RNP-2) and $\beta 3$ (RNP-1) adjacent. Other conserved residues are part of the hydrophobic core (including I, L, or V at 12 and 14; L16, L30, F34; I, L, or V at 40 and 45; V57, F59, A66) or in the capping box of the $\alpha 1$ helix (G38) and less so G74 of $\alpha 2$, using the numbering scheme of RBD1 from the human U1A protein.

Despite the accumulating data that describe this protein and its functions, several fundamental questions about this domain are unresolved. Although the members of this protein family apparently have similar secondary and tertiary structures, they share little sequence identity/homology (Birney et al., 1993). The

Reprint requests to: Kathleen B. Hall, Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110; e-mail: hall@bionmr3.wustl.edu.

Abbreviations: RBD, RNA binding domain; RBD1, N-terminal RBD of U1A (amino acids 1–95); RRM, RNA recognition motif; RNP, ribonucleoprotein; ^{19}F -Tyr, m- ^{19}F fluoro-DL-tyrosine; GdnHCl, guanidine hydrochloride; HOESY, heteronuclear Overhauser exchange spectroscopy; TFA, trifluoroacetic acid.

Fig. 1. Structure of the N-terminal RBD of the U1A protein, with the tyrosine side chains indicated. (From the reconstruction of van Gelder et al. [1994], using the structure of Nagai et al. [1990].) The numbering scheme is shown with the secondary structure elements indicated.

identification of those amino acids that are critical for structural integrity and protein stability and those that control RNA recognition could contribute to an accurate prediction of RBD function. In addition, this information will allow a more direct manipulation of protein structure and/or RNA recognition. Because structural data of the human U1A N-terminal RBD are available as well as some biochemical data, it is used here as a model system to measure the effect of selected mutation on protein stability, structure, and RNA association.

The U1A protein is a part of the U1 snRNP (small nuclear RiboNucleoprotein Particle), where it binds to stemloop II of the U1 snRNA (Scherly et al., 1989; Lutz-Freyermuth et al., 1990). U1A has two RBDs; only the N-terminal domain is required for RNA binding (Scherly et al., 1989; Lutz-Freyermuth et al., 1990; Nagai et al., 1990; Jessen et al., 1991; Hall & Stump, 1992). This domain is small (about 90 amino acids), soluble, and has been crystallized both free (Nagai et al., 1990) and bound to a small RNA hairpin (Oubridge et al., 1994). Many mutations have been made in this domain (Scherly et al., 1989, 1990; Boelens et al., 1991; Lutz-Freyermuth et al., 1990; Nagai et al., 1990; Jessen et al., 1991), usually to determine the effect on RNA binding. There are several biochemical investigations of the RNA:protein interaction that further define this functional aspect of the domain (Jessen et al., 1991; Hall & Stump, 1992; van Gelder et al., 1993; Hall, 1994; Hall & Kranz, 1995; Stump & Hall, 1995).

In this series of experiments with RBD1 of U1A, the four tyrosine residues in the domain have been replaced individually with phenylalanine, and with one other selected amino acid. The

tyrosines were chosen for several reasons: first, because their distribution in RBD1 allows investigation of the different structural elements of the protein (Fig. 1). Tyrosine 13 is on $\beta 1$, in the middle of the four-stranded antiparallel β -sheet; tyrosine 31 is in the middle of $\alpha 1$ on the other side of the sandwich; tyrosine 78 is in the apparently flexible loop 5 between $\alpha 2$ and $\beta 4$; and tyrosine 86 is at the end of $\beta 4$, near the flexible carboxyl terminal tail of the domain. Second, tyrosine 13 is highly conserved among RBDs (Birney et al., 1993), and is known to be involved in RNA binding (Merrill et al., 1988; Oubridge et al., 1994; Stump & Hall, 1995); the other tyrosines are unique to RBD1. Third, m-fluoro-tyrosine could be incorporated into the protein for use in ^{19}F -NMR experiments, which allow observation of the local environment of the tyrosine side chain. Measurements of protein stability, RNA affinity, and protein structure could therefore be combined to derive more information about both the organization and the function of these domains.

Results

The N-terminal domain of the U1A protein is sufficient for high-affinity RNA recognition (Scherly et al., 1989; Lutz-Freyermuth et al., 1990; Nagai et al., 1990; Jessen et al., 1991; Hall & Stump, 1992; Oubridge et al., 1994). The construct used in these experiments, the 1–95 amino acid domain, also binds specifically to the RNA hairpin, but does so with an affinity approximately 30-fold weaker than the 1–102 amino acid domain (Hall, 1994; Stump & Hall, 1995). This shorter domain was used in experiments to determine the effect of mutation at each of the four

tyrosines in RBD1: Y13, Y31, Y78, and Y86. The effects of these substitutions on RBD stability as well as RNA affinity are measured by chemical denaturation and nitrocellulose filter binding, respectively. In addition, ^{19}F -Tyr-labeled RBD proteins are used in NMR experiments to investigate the local environment surrounding each tyrosine as each residue is substituted with phenylalanine and also when RNA is bound.

RBD stability

Each of the four tyrosines in RBD1 has been substituted individually to give eight new RBD1 constructs: Y13F, Y13T, Y31F, Y31S, Y78F, Y78D, Y86F, and Y86T mutant RBD1 proteins. Positions of each tyrosine in the tertiary structure of RBD1 are shown schematically in Figure 1. The standard unfolding free energies, ΔG_D° , for wild-type RBD1 and all mutant proteins were determined by chemical denaturation. Because the CD spectra of the variants were identical to that of the wild-type protein (Hall & Stump, 1992; data not shown), CD was used to measure protein denaturation. The ellipticity at 223 nm (θ_{223}) of each protein was measured as a function of GdnHCl concentration; Figure 3 shows a representative data set for unfolding of wild-type RBD1. Solid lines represent nonlinear least-squares best fit to the linear extrapolation method. In all cases, unfolding was assumed to proceed through a two-state mechanism, although this assumption has not been verified directly. The errors associated with each free energy fall in the range of 0.5–1.0 kcal/mol, and are due mainly to long extrapolation to 0 M GdnHCl.

The unfolding free energy changes for all RBD1 protein variants are given in Table 1, as well as the difference in unfolding free energies, $\Delta\Delta G_D^\circ$, calculated as [$\Delta G_D^\circ(\text{mutant}) - \Delta G_D^\circ(\text{wild type})$]. Transition midpoints range from 3.8 M GdnHCl for wild-type RBD1 to 2 M GdnHCl for the Y78D mutant. The effect of mutations on RBD1 stability can be grouped into three classes. The phenylalanine substitutions at positions 31 on $\alpha 1$

and 78 in loop 5 (between $\alpha 2$ and $\beta 4$) have no destabilizing effects within experimental error; the midpoints and slopes of the unfolding transitions of these proteins do not deviate significantly from that of wild-type RBD1. In the second group of mutations, Y13F and Y13T on $\beta 1$, Y31S, and Y78D destabilize the protein. The effect of the Y13F mutation is small but measurable, and that of the Y13T mutation is greater, destabilizing the RBD1 domain by 1.3 kcal/mol. The Y31S mutation is destabilizing by 2.3 kcal/mol. The Y78D substitution has the most severe effect on the RBD1 domain, reducing protein stability by 5 kcal/mol. The large error associated with the measured stability of Y78D comes from low confidence in the native baseline function, with the unfolding transition beginning at approximately 1.0 M GdnHCl. The final group of mutations, those at position 86 at the end of $\beta 4$, are destabilizing independent of the specific substitutions. Both the Y86F and Y86T mutations cost the domain at least 2.5 kcal/mol of unfolding free energy.

RNA binding affinity

One of the RNA substrates of the N-terminal RBD of U1A is a hairpin, stemloop II of U1 snRNA. A short form of this RNA is used for binding experiments (Fig. 2A). Table 2 gives dissociation constants and the standard free energy (ΔG°) of binding this RNA to the RBDs. In these conditions (100 mM NaCl, 10 mM sodium cacodylate, pH 7, 22 °C), this shortened protein domain (1–95 amino acids) binds the wild-type RNA with $\Delta G^\circ = -13.3$ kcal/mol. Difference free energies ($\Delta\Delta G^\circ$) are relative to this value.

Mutation at Y13 results in a large loss of RNA binding affinity; mutations of the other tyrosines, with one exception, do not reduce affinity for the RNA. Substitution of Y13 with phenylalanine reduces the affinity for the RNA by about 1,000-fold, with a loss of 3.6 kcal/mol of binding free energy. Both struc-

Table 1. Standard unfolding free energies of wild-type and mutant RBDs

RBD1	$\Delta\Delta G_D^\circ$ (kcal/mol) ^a	$\Delta\Delta G_D^\circ$ ^b
Wild type	8.1 (± 0.7)	0
^{19}F wild type	8.1 (± 0.7)	0
Y13F	7.4 (± 0.6)	-0.7
Y13T	6.8 (± 0.6)	-1.3
Y31F	8.1 (± 0.5)	0
Y31S	5.8 (± 0.4)	-2.3
Y78F	8.2 (± 0.5)	0
Y78D	3.0 (± 1.0)	-5.1
Y86F	5.5 (± 0.8)	-2.6
Y86T	5.2 (± 0.6)	-2.9

^a ΔG_D° is the standard unfolding free energy in the absence of GdnHCl, determined by the linear extrapolation method as described in the Materials and methods.

^b $\Delta\Delta G_D^\circ$ is the difference in unfolding free energy referenced to wild-type RBD1, calculated as [$\Delta G_D^\circ(\text{mutant}) - \Delta G_D^\circ(\text{wild type})$]. Unfolding free energies were measured in 50 mM NaCl, 10 mM sodium cacodylate, pH 7.0, at 22 °C.

Fig. 2. Hairpin sequences of SP6 RNA transcripts used in these experiments, which vary in stem length and sequence. **A:** 26-nt RNA used in binding experiments. **B:** 20-nt RNA used in NMR experiments.

Table 2. Binding of proteins to wild-type RNA hairpin^a

RBD1	K_D (M)	ΔG° (kcal/mol)	$\Delta\Delta G^\circ$
Wild-type	2.0×10^{-10}	$-13.1 (\pm 0.1)$	0.0
Y13F	1.0×10^{-7}	$-9.5 (\pm 0.2)$	3.6
Y13T	5.0×10^{-7}	$-8.5 (\pm 0.1)$	4.6
Y31F	3.0×10^{-10}	$-12.9 (\pm 0.2)$	0.0
Y31S	3.0×10^{-10}	$-12.9 (\pm 0.2)$	0.0
Y78F	2.0×10^{-10}	$-13.5 (\pm 0.2)$	0.0
Y78D	5.0×10^{-9}	$-11.2 (\pm 0.2)$	1.9
Y86F	2.0×10^{-10}	$-13.1 (\pm 0.2)$	0.0
Y86T	1.0×10^{-9}	$-12.1 (\pm 0.2)$	1.0

^a The 26-nt RNA hairpin (Fig. 2A) was used in all binding experiments. Binding experiments were performed in 100 mM NaCl, 10 mM sodium cacodylate, 1 mM MgCl₂, pH 7.0, at 22 °C. $\Delta\Delta G^\circ$ is the relative change in binding free energy caused by tyrosine substitutions, calculated as [ΔG° (mutant) – ΔG° (wild type)].

tural (Oubridge et al., 1994) and biochemical (Stump & Hall, 1995) data show that the aromatic tyrosine ring stacks over cytosine 13 (see Fig. 2A) in the RNA loop. Phenylalanine 13 in the Y13F mutant is also presumed to stack, as indicated by its efficient crosslinking to the RNA (Stump & Hall, 1995). Although the Y13F substitution can maintain the stacking of the aromatic side chain with the RNA base, the large loss of RNA binding affinity of the RBD suggests that other RNA:protein contacts are lost. As shown here, the Y13T mutation further reduces affinity, suggesting that the short polar threonine side chain interacts in a different way with the RNA.

One other substitution, Y78D, has an effect on RNA binding, reducing the affinity by 25-fold. This mutation also greatly reduces the stability of the protein, and it seems likely that the protein structure has been disrupted, which may explain the loss of RNA affinity. With the possible exception of this variant, it seems clear that for these mutants there is no correlation between protein stability and RNA affinity.

¹⁹F-Tyr RBD1

Each of the four tyrosines in RBD1 can be monitored using ¹⁹F-tyrosine in NMR experiments that observe directly the ¹⁹F nucleus. The fluorine nucleus is naturally abundant, yet normally absent from most biological molecules and buffers, so it can be introduced in specific residues in the protein; large quantities of these RBDs with nearly 95% incorporation of ¹⁹F-tyrosine can be purified (from mass spectrometry, data not shown). Direct NMR observation of ¹⁹F is facilitated by its efficient detection (83% of ¹H sensitivity), and its sensitivity to its chemical and magnetic environment (Gerig, 1978). For these experiments, the major problem with ¹⁹F detection is that the size of the protein and the field strength (470 Mz) with the resulting chemical shift anisotropy contribution to the relaxation (Hull & Sykes, 1975) combine to produce broad resonances in the NMR spectrum.

Tyrosine substitution with the electronegative fluorine could possibly affect protein stability or RNA binding, by altering the polarizability or chemistry of the ring (the p*K_a* of fluoro-tyr is estimated between 8.3–9.5 [Dettman et al., 1985; Carne et al.,

1986], whereas that of normal tyr is 10.3) and so change the hydrogen bonding properties of the hydroxyl group. To determine the effect of fluoro-tyrosine substitution on the protein, GdnHCl denaturation of Y13F, Y31F, Y78F, and Y86F proteins containing fluoro-tyrosine was used to measure the unfolding free energy. As shown in Figure 3, the unfolding of fluoro-Tyr RBD1 is identical to unfolding of the unsubstituted protein, within experimental error (Table 1); no effect on protein stability was observed for any fluoro-RBD. Similarly, introduction of the fluoro-tyrosine did not alter RNA binding properties, as determined by nitrocellulose filter binding assays (data not shown). Therefore, introduction of fluoro-tyrosine does not alter protein stability or protein function.

1D ¹⁹F NMR

Individual ¹⁹F-tyrosine resonances in the NMR spectrum were assigned by direct comparison of RBDs containing single tyrosine-to-phenylalanine substitutions. Shown in Figure 4A are the 1D spectra of the wild-type RBD1, the Y13F, Y31F, and Y86F proteins. As seen from these spectra, the single mutations do not alter the chemical shifts of the remaining ¹⁹F-tyrosine resonances, making assignment unambiguous. In these spectra (25 °C, 2 mM protein), the signal from Y78, which is near that of Y86, is significantly broadened and is not distinguishable from the baseline, as is best illustrated in the spectrum of the Y86F mutant. Chemical shifts of the resonances in ¹H₂O (²H₂O) are Y13: –60.91 (–60.97) ppm; Y86: –59.96 (–59.99) ppm; and Y31: –61.11 (–61.21) ppm.

The temperature dependence of wild-type ¹⁹F-tyr RBD1 is shown in Figure 4B; at 40 °C, the Y78 resonance becomes observable as a broad peak downfield of the Y86 resonance. The ¹⁹F chemical shifts of Y13, Y31, and Y86 vary linearly with temperature from 4 to 50 °C, at which point the protein aggregates. Y13 exhibits the largest temperature-dependent chemical shift changes, and Y86 the smallest. At low temperature, the Y31 and Y13 resonances overlap. The appearance of the Y78 reso-

Fig. 3. Ellipticity at 223 nm as a function of GdnHCl concentration for wild-type RBD1 (open circles) and for ¹⁹F-Tyr-labeled wild-type RBD1 (closed circles) in the presence of 50 mM NaCl and 10 mM sodium cacodylate at pH 7.0 and 23 °C. Solid lines are the nonlinear least-squares best fit to the data using the linear extrapolation model. Data sets used for calculation of ΔG° are averages of duplicate experiments with 30 separate denaturant concentrations; standard deviation for each data point does not exceed 4%.

Fig. 4. 1D ^{19}F spectra of proteins. **A:** 1D ^{19}F spectra of ^{19}F -Tyr RBD1 proteins, 25 °C: wild-type RBD1, Y31F RBD1, Y13F RBD1, and Y86F RBD1. The Y78 resonance is not observed at this temperature. All proteins are 2 mM. **B:** Temperature-dependent 1D ^{19}F spectra of 2 mM ^{19}F -Tyr RBD1. At 40 °C, the Y78 resonance is observed as a broad peak to the side of the Y86 resonance. **C:** 1D ^{19}F spectra of wild-type RBD1 at 0.4 mM. The Y78 resonance is now observed. All solutions contained 100 mM NaCl, 10 mM NaPO₄, 1 mM MgCl₂, pH 6.0 (8% D₂O).

nance at high temperature may indicate that the tyrosine side-chain motion is in intermediate exchange on the NMR timescale, and therefore its signal is motionally broadened.

Tyrosine 78 is in the long hydrophobic loop 5 (amino acids 73–81) between $\alpha 2$ and $\beta 4$. The behavior of this region of the protein is unusual in solution. As these experiments show, the ^{19}F resonance on the aromatic side chain is not observed in these spectra below about 40 °C (although at pH 8.8 in borate, it is present as a broad resonance), suggesting that some conformational averaging is occurring. $^1\text{H}/^{15}\text{N}$ HSQC experiments of 1.5 mM 1–102 amino acid protein show that the F77, D79, K80, and M82 backbone amides give very weak signals

at 25 °C, but at 30 °C, their crosspeaks become much stronger (see Fig. 5). The Y78 amide is apparently normal, although its crosspeak is in a crowded region of the spectrum. The dynamics of this region of the protein differ from those of the surrounding residues, as the T1, T2, and NOE data for the amide $^1\text{H}/^{15}\text{N}$ show (Fig. 6). The T1 value of the F77 amide is significantly longer than that of the surrounding residues, and the T2 values of F77 and D79 are significantly shorter. The dynamics measurements of backbone amides and the ^{19}F data of the Y78 side chain both suggest that this loop of the protein is undergoing some slow conformational averaging on the μs to ms timescale.

Fig. 5. $^1\text{H}/^{15}\text{N}$ HSQC data of the 1–102 amino acid domain at (A) 25 °C and (B) 30 °C. Crosspeaks are labeled with the one-letter amino acid codes and residue numbers. Vertical lines connect crosspeaks from NH_2 groups of asparagine and glutamine side chains. $\text{NH}\epsilon$ protons of arginine are folded in this experiment and labeled 91–95. Temperature-dependent residues from loop 5 are indicated in B (2 mM protein, 100 mM KCl, 30 mM KPO₄, pH 5.5).

Fig. 6. Relaxation data for RBD1 as a function of residue number. ^{15}N T1, T2, and NOE values measured at 500 MHz, 25 °C, 2 mM protein, 100 mM KCl, 30 mM KPO_4 , pH 5.5.

However, the solution behavior of loop 5 is concentration dependent. At 0.5 mM protein, the ^{19}F resonance from Y78 can be seen at -60.2 ppm in the spectrum of the wild-type RBD1 (Fig. 4C). $^1\text{H}/^{15}\text{N}$ HSQC spectra of the wild-type protein at

this concentration show nearly normal behavior of all residues in loop 5, although the T2 values of the F77 and D79 amides are slightly shorter than the average amide value (J. Lu & K.B. Hall, in prep.). Inelastic light scattering measurements of RBD1

at 1.5 mM showed it to be polydisperse; at 0.1 mM, analytical ultracentrifugation experiments showed it to be monomeric (data not shown). Finally, after addition of 0.2 M GdnHCl to the 2 mM RBD1 sample, the Y78 resonance could be observed, although it was still somewhat broader than the other peaks (data not shown). Together, all these data indicate that, at high concentrations, the loop 5 residues are participating in protein self-association.

HOESY experiments with ^{19}F -Tyr RBD1

In order to probe the structure of the regions surrounding each tyrosine in more detail, 2D HOESY was used to look for protons in proximity to the ^{19}F nucleus. One potential advantage of this experiment is that, because the ^{19}F is observed in the acquisition dimension, the spectra can be recorded in $^1\text{H}_2\text{O}$, because there is no contribution of the water resonance to overload the signal. One disadvantage is that the ^{19}F resonances are broad, making the spectra rather misshapen. However, any protons within 5 Å of ^{19}F on the tyrosine should be detectable in this experiment, and thus it provides a means to probe the local environment at each site.

Wild-type RBD1

HOESY spectra of the ^{19}F -Tyr RBD1 were acquired in $^1\text{H}_2\text{O}$ and in $^2\text{H}_2\text{O}$. Examples of both spectra are shown in Figure 7. The most notable crosspeaks at each ^{19}F -Tyr resonance come from protons at 4.8 ppm. These crosspeaks are absent in the spectra collected in $^2\text{H}_2\text{O}$, and are therefore assigned to HOEs from water protons. The HOE from $^1\text{H}_2\text{O}$ is the most pronounced crosspeak observed at short mixing times (50 ms), indicating that these tyrosines are solvent accessible (Cistola & Hall, 1995). The intensity of other crosspeaks is enhanced in the $^2\text{H}_2\text{O}$ spectrum, suggesting that magnetization transfer to water is a dominant relaxation pathway for these ^{19}F nuclei. For this reason, most other experiments were done in $^2\text{H}_2\text{O}$.

To investigate these $^1\text{H}_2\text{O}$ - ^{19}F HOE crosspeaks, and to determine the chemistry of the tyrosines in the protein, HOESY experiments were done with the free m- ^{19}F -tyrosine amino acid, as well as m- ^{19}F -DL-phenylalanine. Both amino acids showed strong crosspeaks from water to the attached ^{19}F nucleus (data not shown), indicating that this observation is not dependent on the presence of the tyrosine hydroxyl group. The pK_a of each tyrosine residue in the protein was also measured using the change in ^{19}F chemical shift as a function of pH, to ascertain the ionization state of the side chain. The magnitude of the chemical shift change varied with the residue, with Y86 showing the biggest changes and Y13 the smallest. However, the titration indicated that the pK_a of each is approximately 9.0 ± 0.1 (data not shown). This value is below the pK_a of tyrosine, but apparently typical of the fluorinated amino acid, indicating that all tyrosine hydroxyl groups are in the protonated form at the pH of the NMR experiments.

In addition to the crosspeak from water, each ^{19}F resonance shows several other HOE crosspeaks from protons on the protein. The closest proton will be the adjacent $^1\text{H}_2$ on the tyrosine ring; in the free amino acid in these solution conditions, the $^1\text{H}_2$ - ^{19}F crosspeak occurs at 7.21 ppm. For Y31, that proton can be confidently assigned to the crosspeak at 7.22 ppm; for Y86 it appears at 7.31 ppm; and for Y13, that proton is at 6.31 ppm. No crosspeak was observed to Y78 at 1.5 mM

[RBD1], but its adjacent $^1\text{H}_2$ can be observed at 7.3 ppm in an experiment with 0.4 mM protein (data not shown). By reference to the free amino acid, Y13 appears to be in a unique environment. The upfield shift of Y13 $^1\text{H}_2$ may be evidence that tyrosine 13 is stacked with another aromatic ring; F56 on the adjacent $\beta 3$ strand would be the most likely candidate.

There are additional protons that are close to these ^{19}F -tyrosines, but few are assigned here. A weak crosspeak from an aromatic proton at 7.3 ppm to Y13 is observed in the HOESY in $^2\text{H}_2\text{O}$. There are crosspeaks from aliphatic protons (1–3 ppm), and a strong HOE to Y31 from a proton at 0.3 ppm, which is likely a methyl group. There are several strong HOE crosspeaks to Y86, particularly in the region from 1 to 2 ppm (^1H) from aliphatic protons.

With these data, it is possible to compare HOESY spectra of mutant proteins, to look for changes in the HOE patterns at each site. In particular, this comparison of local environments can be used to determine if a mutation at one site on the protein can perturb the conformation at other positions, through propagation of structural changes.

^{19}F -Tyr Y13F RBD

Tyrosine 13 is in the $\beta 1$ strand, one of two conserved sequences among RBD proteins (Birney et al., 1993). In the crystal structure (Nagai et al., 1990), there appears to be a hydrogen bond between the Y13 hydroxyl and the side-chain carbonyl oxygen of glutamine 54. The Y13F mutation would lose this hydrogen bond, which is perhaps the cause of its slightly reduced standard unfolding free energy ($\Delta\Delta G^\circ = -0.7$ kcal/mol). However, as shown in Figure 8A, the HOESY spectrum of this protein is virtually identical to that of the wild-type protein.

^{19}F -Tyr Y31F RBD

Tyr 31 is in the middle of the $\alpha 1$ helix facing the protein exterior; its ^{19}F has a strong HOE from $^1\text{H}_2\text{O}$. Substitution of phenylalanine for tyrosine is isoenergetic; not surprisingly, the Y13 and Y86 resonances show essentially the same HOE patterns as in the wild-type RBD1 spectrum (Fig. 8B).

^{19}F -Tyr Y78F RBD1

Tyrosine 78 is in the hydrophobic greasy loop 5 between $\alpha 2$ and $\beta 4$. The HOESY spectrum of Y78F RBD (Fig. 8C) suggests that this substitution has altered the protein. Although the HOE crosspeaks to ^{19}F -Tyr 31 are identical to those for the wild-type protein, there are additional crosspeaks to both Y13 and to Y86 in this RBD. As in the wild-type protein, there is a crosspeak to ^{19}F -Y13 from a proton at 7.3 ppm, but its intensity is significantly increased in this mutant. There are two new crosspeaks centered around 5.2 ppm, in addition to new crosspeaks at 0.6 ppm and 1.3 ppm. At Y86, which is on the opposite end of the $\beta 4$ strand, there are several new crosspeaks from protons at 0–2 ppm, and two near 3.3 ppm. It is also interesting to note that the line width of the Y13 resonance has substantially increased to about 70 Hz, and that its proton crosspeaks do not occur at the same fluorine frequency.

These results suggest that the phenylalanine substitution in loop 5 has caused some change in the environment around Y86 and Y13, but not at Y31. All crosspeak intensities are increased relative to the wild-type protein. Because loop 5 is implicated in protein:protein contacts, it may be that, in this mutant at this

Fig. 7. Comparison of ^{19}F -detected HOESY spectra of wild-type ^{19}F -Tyr-labeled RBD1 in 92% H_2O /8% D_2O (spectrum 7B) and in 100% D_2O (spectrum 7A) at 25 $^\circ\text{C}$. All resonances are referenced to TFA at 0 ppm in 92% H_2O /8% D_2O . These HOESY spectra were acquired with a 300-ms mixing time, but the strong HOE from $^1\text{H}_2\text{O}$ is observed also at shorter mixing times (50 ms), where it can be the strongest crosspeak. Data were weighted with a shifted gaussian and zero-filled to 2,048 points in both dimensions (2 mM protein in 100 mM NaCl, 10 mM NaPO_4 , 1 mM MgCl_2 , pH 6.0).

protein concentration, protein association is reduced and therefore HOE efficiency is greater.

^{19}F -Tyr Y86F RBD

Tyrosine 86, at the end of the last β strand (β_4), is near the flexible carboxyl tail of the RBD. The HOESY spectrum (Fig. 8D)

shows that the crosspeaks from protons to ^{19}F -Y31 are identical to those in the wild-type protein. However, three new crosspeaks are found to ^{19}F -Y13. Again in this mutant, the ^{19}F -Y13 resonance is broad; two peaks can be resolved from the 2D spectrum, separated by approximately 29 Hz. Their combined line width is about 60 Hz.

Fig. 8. HOESY spectra of ^{19}F -Tyr-labeled mutant RBD1 proteins, in 100% D_2O at 25 $^{\circ}\text{C}$. Mutants are (A) Y13F, (B) Y31F, (C) Y78F, and (D) Y86F. Data were processed as in Figure 7, and collected in the same solution conditions.

Orientations of the tyrosine ring

Although the fluorine resonances are very broad in these spectra, it is clear that there are ^{19}F resonances that appear to be composed of two peaks. One explanation is that there are two (nearly) distinguishable orientations of the ^{19}F . In the case of the Y13 side chain, the hydrogen bond from Y13 to Q54 may constrain the rotation about the $\text{C}\beta\text{-C}\gamma$ bond of the tyrosine, slowing it down so that the asymmetrically positioned ^{19}F nucleus is able to report on orientations of the ring. The HOESY spectrum shows evidence of two distinct environments of this ^{19}F for the Y86F RBD, and perhaps for the wild-type and Y78F proteins. The separation between the two discernable peaks in the Y13 resonance is 29 Hz in the Y86F mutant, and 19 Hz in the wild-type protein at 25 °C (measured from the tops of the peaks in the HOESY spectra; note that in 1D, these peaks are unresolved). At 40 °C, the separation of the two ^{19}F -Y13 peaks measured in a HOESY experiment with the Y86F mutant decreases, consistent with the interpretation that the tyrosine side chain exists in two stable orientations that are in intermediate exchange in these conditions. As predicted for this interpretation, at higher temperatures, the rate of exchange increases and the two resonances coalesce. The rate of interconversion is not estimated here as the large ^{19}F line width makes measurement too imprecise. The ^{19}F -Y13 resonance in the spectrum of the Y78F RBD is very broad, but does not show obvious evidence of two distinct frequencies of the ^{19}F .

^{19}F -RBD1:RNA complexes

This domain of the human U1A protein interacts with an RNA hairpin, either in the context of the entire U1 snRNA, or as a separate RNA fragment. For these experiments, the RNA hairpin shown in Figure 2B is bound to the wild-type RBD1, such that all the protein is present as a complex. The subnanomolar association constant assures that only the bound species is observed. The RNA sits on the surface of the β -sheet, where it makes hydrogen bonds with many amino acid side chains and

backbone amide and carbonyl groups (Oubridge et al., 1994). Of the four tyrosine residues, only Y13 is close to the RNA; Y13 is stacked over the C13 base in the RNA loop and the two can be crosslinked efficiently (Stump & Hall, 1995). It was anticipated that its ^{19}F chemical shift would be changed in the complex; no other tyrosine is directly involved with RNA association.

1D spectra of ^{19}F -RBD1:RNA

Changes in the 1D ^{19}F spectrum of the RBD1:RNA complex are shown in Figure 9. As a consequence of the increase in molecular weight of the complex (17 kDa), all fluorine line widths have increased about 35%. More interesting, two ^{19}F resonance changes are observed: ^{19}F -Y13 is shifted upfield from -60.97 to -61.09 ppm (25 °C), and ^{19}F -Y78 can now be seen at lower temperatures. Other resonances do not change. These results are surprising, first because a larger change in the chemical shift of Y13 was anticipated given its position with respect to the RNA base, and second because the appearance of Y78 suggests that this side chain now has different motional properties. The likely explanation for the latter observation is that binding of the RNA prevents intermolecular association so that the Y78 side chain in the complex has the dynamic characteristics of the monomeric protein.

HOESY spectra of the complex were complicated by the faster relaxation rates of the ^{19}F nuclei, and mixing times were considerably shortened. Also, HOESY experiments of the complex in $^1\text{H}_2\text{O}$ yielded no unambiguous proton crosspeaks, except a strong HOE from the solvent $^1\text{H}_2\text{O}$ to all resonances (data not shown). This result was especially interesting for Y13, and shows that, even though it is stacked over an aromatic nucleotide base, it is still solvent accessible. A HOESY spectrum of the complex in $^2\text{H}_2\text{O}$ is shown in Figure 10. At this mixing time (175 ms), the HOE crosspeaks from protons in the free protein are reduced in number and intensity from those of Figure 7.

Although the crosspeak intensities in the HOESY spectrum of the complex are weak, it is clear that the environments around Y13 and Y86 are altered as a consequence of binding the RNA.

Fig. 9. 1D ^{19}F spectra of ^{19}F -Tyr-labeled RBD1 bound to the 20-nt hairpin RNA as a function of temperature. Assignment of ^{19}F -Tyr peaks was determined from complexes of RNA and RBD1 variants.

Fig. 10. HOESY spectrum of RNA: ^{19}F -Tyr RBD1 complex at 25 °C in $^2\text{H}_2\text{O}$. Processing parameters as in Figure 7, with a 175-ms mixing time. The complex concentration was 1 mM in 100 mM NaCl, 10 mM NaPO_4 , 1 mM MgCl_2 , pH 6.0.

There is one new proton crosspeak to Y13 at 2.3 ppm: its own $^1\text{H}_2$ is slightly shifted to 6.2 ppm, and there is a weak crosspeak near 7.3 ppm, which is not seen in the protein at this mixing time. It is somewhat puzzling that even though an RNA base (C13) is within 5 Å of Y13, its chemical shifts and those of its proximal neighbors are nearly unchanged. Equally puzzling is the absence of proton crosspeaks from the RNA. The environment of Y86 is also changed in the complex: there is one strong new crosspeak near 0.8 ppm, but one at 2.1 is missing. It appears, from this limited sampling of the protein, that the chemical environment of its β -sheet surface and of peripheral loops has been altered in response to RNA binding.

Discussion

There are four tyrosine residues in the human U1A N-terminal RBD. In these experiments, each has been substituted individually with phenylalanine and one other amino acid, to determine their contribution to protein stability and to RNA binding. Of the four tyrosines, only Y13 in the β 1 strand is conserved among RBDs. It is located in the RNP-2 consensus sequence of the protein and is part of the RNA binding surface. The other three tyrosines, Y31 in the α 1 helix, Y78 in loop 5, and Y86 at the end of β 4, are peculiar to this domain. These results show that each side chain contributes in a unique way to the structure and function of this protein.

Tyrosine 13

Tyrosine/phenylalanine at position 13 in β 1 (41% Phe/31% Tyr) and phenylalanine at position 56 in β 3 (70% Phe/7% Tyr)

are highly conserved aromatic residues in all RBDs (using the alignment of Birney et al., 1993). The side chains of these two aromatic residues point into solution in all known RBD structures (human U1A RBD1, Nagai et al., 1990; hnRNP C1, Gorlach et al., 1992; Wittekind et al., 1992; sex lethal RBD2, Lee et al., 1994; U1A RBD2, J. Lu & K.B. Hall, in prep.). Phe 16 in hnRNP A1 is at the same relative position as Tyr 13 in U1A RBD1; Phe 16 was covalently photo-crosslinked to dT₈ (Merrill et al., 1988), and Tyr 13/Phe 13 were covalently cross-linked to 5-I-U13 of the RNA hairpin (see Fig. 2A) (Stump & Hall, 1995). In the cocrystal of RBD1 with an RNA hairpin, Tyr 13 stacks over C13 (Oubridge et al., 1994), and in the ^{19}F NMR experiments, ^{19}F -Tyr 13 undergoes a significant change in chemical shift relative to other ^{19}F resonances when the RNA binds. It seems likely that in all RBD:RNA complexes these aromatic groups provide a nonspecific stacking interaction with the RNA bases.

As these experiments show, Tyr 13 can be replaced with phenylalanine or with threonine with only slight loss of protein stability: the loss of 0.7–1.3 kcal/mol of unfolding free energy is small but significant. In the crystal structure of RBD1, the hydroxyl group of Y13 makes a hydrogen bond to the side-chain carbonyl oxygen of Q54 (Oubridge et al., 1994). This interaction will be absent in the F13 mutant, and the threonine side chain is not long enough to maintain this contact. In theory, tyrosine 13 is close enough to phenylalanine 56 for the two aromatic rings to stack, which would provide a favorable hydrophobic interaction. This interaction, if it is present, could be maintained in the F13 substitution, but absent in the T13 variant. Given the network of hydrogen bonds that link Y13 to Q54 to the amino acid side chains of loop 3 (Oubridge et al., 1994), it is too simplistic to assign the 1 kcal/mol of unfolding free energy to the

Y13/Q54 hydrogen bond and loss of a possible stacking interaction between Y13 and F56. However, the magnitude of the destabilization observed for the two mutant RBDs is consistent with the loss of these interactions.

Both the F13 and T13 variants bind only weakly to the RNA hairpin. In the case of the F13 mutation, there is a loss of 3.6 kcal/mol of binding free energy, and 4.6 kcal/mol for the T13 mutation upon formation of the RNA:RBD complex. Because phenylalanine at position 13 can be crosslinked to the RNA, it seems clear that it adopts the same orientation with respect to the RNA base. The loss of binding affinity must therefore come from the loss of other RNA:protein contacts. In the cocrystal, Y13 maintains a hydrogen bond to Q54 (Oubridge et al., 1994) and the side-chain amino protons of glutamine 54 are also able to hydrogen bond to the side-chain carbonyl oxygens of R52 and K50 in loop 3 between $\beta 2$ and $\beta 3$. Dynamics measurements of the protein show that several residues in this loop undergo conformational exchange (J. Lu & K.B. Hall, in prep.), suggesting that it is relatively flexible. In the complex with RNA, this loop makes contact with the top of the RNA stem at the base of the RNA loop, and participates in a network of interresidue protein hydrogen bonds (Oubridge et al., 1994). One hypothesis for the role of the unique Q54 in RBD1 is that it may be directly involved in stabilizing the backbone geometry of loop 3 to facilitate these interactions. The large reduction in RNA affinity of the F13 mutant may result from an increase in the conformational freedom (entropy) of the glutamine 54 side chain that no longer anchors the loop 3 backbone. The further reduction in binding free energy for the T13 variant could be attributed to the additional loss of the stacking interaction between nucleotide base and aromatic amino acid side chain.

The NMR data provide support for the presence of these interactions. In the HOESY spectra, there are crosspeaks to ^{19}F -Y13 from its own H2 proton at 6.3 ppm, and from one other aromatic proton at 7.3 ppm. This latter proton is tentatively assigned to F56, which has its C ϵ within 3.5–4.2 Å of the Y13 C ϵ [from the reconstruction of the RBD structure (van Gelder et al., 1994)]. This juxtaposition is consistent with stacking of these two aromatic side chains. In addition, an interesting feature of the HOESY spectra is the two resonances that are sometimes associated with ^{19}F -Y13. The position of the ^{19}F on the tyrosine ring is asymmetric, but if the C α /C β /C γ bonds were undergoing free rotation that was fast on the NMR time-scale, the orientations of the tyrosine ring would be averaged and one ^{19}F peak would be observed. However, it seems clear that in the HOESY spectra of several of the proteins, it is reporting on two orientations of the side chain. A simple explanation for this observation is that the tyrosine hydroxyl group is hydrogen bonded to glutamine 54, which slows the rotation about the C β /C γ single bond, such that the ^{19}F is held in two distinct orientations. As Hull and Sykes reported (1975), the flip rates around the C β /C γ bond can vary over 3–4 orders of magnitude; clearly, in this example, Y13 is in the limit of slow exchange. This system may be an exciting one for the investigation of the motion of this aromatic ring as a function of solution conditions, temperature, and mutation. However, as the spectra illustrate, the line widths of the ^{19}F at this field strength suffer enormously from the chemical shift anisotropy of this nucleus that dominates T $_2$ relaxation (Hull & Sykes, 1975). Quantitative analysis of relaxation times and motions will require experiments at lower fields.

Tyrosine 31

This residue in $\alpha 1$ was substituted with phenylalanine and serine, to assess the contribution of the aromatic ring and the hydroxyl group. Phenylalanine substitution does not perturb the energetics of the domain, and the HOESY data indicate that this substitution does not alter the environment around the other tyrosine residues. Reciprocally, substitution of the other tyrosine residues with phenylalanine does not change the environment of Y31. Serine substitution was used in part because serine is present at an analogous position in the first α helix of the U1A C-terminal RBD; threonine would have been an alternative substitution to preserve the hydroxyl group. Neither substitution had any effect on RNA binding, which is perhaps not surprising because the α -helices are on the back side of the protein relative to the RNA binding surface. Similarly, Tyr 31 substitution with a histidine (Oubridge et al., 1994) apparently has no effect on RNA affinity. However, replacement of the tyrosine with serine resulted in a measurable reduction in standard unfolding free energy of the RBD.

On the basis of the protein denaturation data, substitution of tyrosine 31 with serine has resulted in the loss of interactions that contribute 2.3 kcal/mol to the protein stability. Although serine is not accommodated in α -helices as well as either tyrosine or phenylalanine, it is unlikely that the destabilization is entirely due to the relative helix-forming propensity of these residues. Using the Zimm–Bragg s value as a measure of helix stability, tyrosine is assigned $s = 1.02$; phenylalanine, $s = 1.09$; and serine, $s = 0.76$ (Wojcik et al., 1990, from host-guest experiments), with the lower value indicating a lower preference for helical conformations. A reasonable value for the energetic effect of the serine substitution in a helix is a reduction in the free energy by 0.3–1.1 kcal/mol (O’Neil & DeGrado, 1990; Blaber et al., 1994; Chakrabarty et al., 1994), which is considerably less than the measured value for the S31 RBD. Perhaps more importantly, relative to the original tyrosine at this position, the energetic penalty for serine occupation is actually minimal (Blaber et al., 1994). A plausible explanation for the observed destabilization is that the tyrosine (as well as the phenylalanine substitution) interacts with the lysine 27 side chain. The spacing of these residues on the same side of the α helix (i to $i + 4$) allows the amino group of the extended lysine side chain to contact the tyrosine aromatic ring, where an interaction between that π electron system and the lysine amino group can contribute to the $\alpha 1$ helix stability. This lys/tyr(phe) interaction may be similar to the charged amino/aromatic interactions suggested first by Burley and Petsko (1986) and discussed in recent work by Mitchell et al. (1994) and Flocco and Mobray (1994), and summarized as general cation- π interactions by Dougherty (1996). The Y31/K27 pair could be another example of a stacked amino group above the plane of the aromatic ring, an interpretation supported by the isoenergetic substitution of F31. If so, then this interaction is contributing substantial free energy to the stability of the domain.

Tyrosine 78

In this RBD, loop 5 between $\alpha 2$ and $\beta 4$ is seven amino acids long, with the sequence PPFY $_78$ DKP. This arrangement of hydrophobic residues is shared only by RBD1 of the highly homol-

ogous U2B" protein (Birney et al., 1993). In most RBDs, this loop contains predominantly polar residues.

In the crystal structure, hydrophobic residues F75, F77, and Y78 from loop 5 are packed tightly against L26, L30, and F34 from $\alpha 1$, forming part of the hydrophobic core of the RBD. In solution, the dynamics data suggest that this region of the domain is undergoing some conformational averaging, possibly related to the presence of two proline residues in addition to variations in intermolecular and intramolecular packing. NMR data suggest that this region is involved in protein:protein association at high protein concentrations, as this greasy patch of the protein could associate with other similarly exposed patches. When RNA is bound, the loop 5 environment around Y78 is altered, probably reflecting disruption of protein association.

Substitution with phenylalanine does not alter the unfolding free energy of the domain, neither does it change the affinity for the RNA. However, the HOESY experiments indicate that the resulting RBD differs from the wild-type protein. This difference may relate to protein:protein association that might be reduced by this mutation; less likely, the packing of loop 5 in the RBD could be different in such a way as to alter the arrangements of side chains on the β -sheet. Introduction of aspartic acid severely destabilizes the protein and reduces affinity for RNA. In general, the energetics data indicate that the packing of loop 5 contributes significantly to the overall stability of the protein.

Tyrosine 86

Tyr 86 is in a region of RBD1 where several structural elements are juxtaposed: it is at the C-terminal end of $\beta 4$, which is packed against $\alpha 2$ residues, along with the N-terminal residues of the conserved $\beta 1$. Tyrosine is not conserved at this juncture, however, as alignment suggests that there is no sequence preference at the end of this last β -strand. Dynamics measurements with the longer 102A protein indicate that Y86 maintains an ordered state; starting from K88, the NOE values become smaller, and T1 and T2 values gradually increase until G99 when they become uniformly small. The last three residues, T100, F101, and V102, have NOE, T1, and T2 values characteristic of a small molecule, indicating that they are largely disordered.

Substitution of Y86 with either F86 or T86 results in a significant reduction of the unfolding free energy of RBD1. The equal loss of folding free energy observed with the F86 and T86 substitutions indicates that the hydrophobic ring is not sufficient to maintain the tyrosine interactions, nor can the hydroxyl group on the short threonine side chain substitute for tyrosine; thus the critical interaction for protein stability is mediated by the tyrosine hydroxyl group. In the reconstruction of the protein side chains (van Gelder et al., 1994), Y86 is tucked between the β -sheet and $\alpha 2$, where its hydroxyl is within hydrogen bonding distance of T66. A probable consequence of this mutation is that the loss of interactions has made this region more dynamic.

HOESY data suggest that the effect of the Y86F mutation has propagated to the environment of Y13. However, binding affinity to the RNA hairpin is not altered by either tyrosine 86 substitution. Several residues surrounding Y86 participate in hydrogen-bonding interactions to the RNA, including backbone carbonyl and amide protons of residues 89–93, and amino acids Y13 and T11 (Oubridge et al., 1994). These interactions are apparently not lost in this moderately destabilized protein.

Conclusions

These experiments with the N-terminal RBD of the human U1A protein were designed to address the contribution of the tyrosines to protein stability, structure, and RNA binding. As this comprehensive dissection indicates, each tyrosine affects the protein in a unique way. This protein is very stable, and although it is small, it has a packed hydrophobic core consisting of many conserved residues. Three of these tyrosines make substantial contributions to its stability. Tyrosine 31 in the $\alpha 1$ helix may interact with lysine 27 to stabilize the helix. Loop 5, which includes tyrosine 78, has unusual solution properties based on its hydrophobic content, and its packing arrangement in the protein contributes significantly to protein stability. Tyrosine 86 certainly contributes to stable packing of the β -strands against the α helices, and may serve to anchor the C-terminal tail of the protein. In contrast, tyrosine 13 has little contribution to the protein stability, but is critical for RNA binding. The network of interactions between Y13, Q54, and loop 3 may determine the structure and dynamics of this region of the protein that control association with the RNA hairpin.

These experiments have also illustrated an important principle about these apparently very similar RNA binding proteins. Although they share an $\alpha\beta$ sandwich global topology, the details of their packing, stability, and possibly their folding pathways, will be unique to each protein. Their selection of RNA targets is certain to be influenced by these features, making their properties as proteins an essential part of understanding and predicting their function.

Materials and methods

Materials

Taq polymerase was obtained from Boehringer Mannheim. L-amino acids, m-Fluoro-DL-tyrosine, and 99% pure GdnHCl were obtained from Sigma Chemical Co. All other chemicals are reagent grade. SP6 RNA polymerase was purified from an over-producing strain of *Escherichia coli* (Jorgensen et al., 1991). DNA oligonucleotides were purified by denaturing PAGE.

RNA synthesis

Details of RNA synthesis are described elsewhere (Stump & Hall, 1993). The wild-type RNA substrate was synthesized from short DNA oligonucleotide substrates using SP6 RNA polymerase. Two separate stemloop RNA molecules were used in these studies, varying in stem length and sequence (Fig. 2); each RNA forms a stable hairpin structure (data not shown). The RNA used in nitrocellulose filter binding assays (Fig. 2A) was labeled with both [α - 32 P]CTP and [α - 32 P]-UTP (NEN) by enzymatic synthesis. All RNAs were purified from polyacrylamide-urea gels.

Site-directed mutagenesis

Mutations in RBD1, defined as the first 95 amino acids of U1A (11 kDa), were generated by site-directed mutagenesis with PCR methods using the wild-type RBD1 clone (Sillikens et al., 1987). Tyrosine mutations at position 13 and 86 (either Y13F, Y13T, or Y86F, Y86T) were made from long PCR primers containing

the mutations. Tyrosine mutations at Y31 and Y78 (either Y31F, Y31S, or Y78F, Y78D) were generated using a recombinant PCR method (Higuchi, 1990). Complete DNA oligonucleotide sequences for all PCR primers are as follows (C-primers are coding strand DNA, N-primers are anti-coding strand DNA): N-terminus, 5'-CGGCCATGGCAGTTCCC; C-terminus, 5'-GGCCAA GCTTCTACTAGGCAATGAT; Y13F, N-5'-CGGCCATGGCA GTTCCCCGAGACCCGCCCTAACCACACTATTTTATC; Y13T, N-5'-CGGCCATGGCAGTTCCCCGAGACCCGCCCTA ACCACACTATTACTATC; Y31F, N-5'-AAGTCCCTGTTCG CCATCTTC and C-5'-TTCAGGGACAAGCGGTAGAAG; Y31S, N-5'-AAGTCCCTGAGCGCCATCTTC and C-5'-TTCA GGGACTCGCGGTAGAAG; Y78F, N-5'-TTCCCTTTCTTTG ACAAACCT and C-5'-AAGGGAAAGAACTGTTTGA; Y78D, N-5'-TTCCCTTTTCGATGACAAACCT and C-5'-AAG GGAAAGCTACTGTTTGA; Y86F, C-5'-GGCCAAGCTTC TACTAGGCAATGATATCTGAGTCGGTCTTGCGAACT GG; Y86T, C-5'-GGCCAAGCTTCTACTAGGCAATGATAT CTGAGTCGGTCTTGCGAGTCTGG. PCR used standard procedures (as described in Sambrook et al., 1989), and products were cloned into a plasmid under control of the tac promoter (Hall & Stump, 1992). All RBD1 mutations were verified by sequencing of the DNA plasmid.

Protein purification

All proteins were overproduced and purified following a standard procedure (Hall & Stump, 1992). Proteins biosynthetically labeled with ^{19}F -Tyr were overproduced in an *E. coli* tyrosine auxotroph (CGSC #4510, strain designation AT2471, from ATCC). Cells were grown in an enriched minimal media also containing 10 $\mu\text{g}/\text{mL}$ thiamine, additional amino acids (5 mM Gln; 500 μM each of Arg, His, Ile, Leu, Lys, Met, Phe, Thr, and Val; 200 μM Trp), and added tyrosine. Protein induction by IPTG (USB) proceeded once the 50 μM tyrosine required to reach mid-log phase was exhausted; addition of IPTG was concomitant with addition of 150 μM ^{19}F -Tyr, followed by four hour induction. In all cases, overproducing cells were lysed by French pressing, followed by protein purification over a CM-Sephare column (Pharmacia). RBDs purified in this manner were >95% pure as assessed by silver stained SDS-polyacrylamide gels. Unlabeled proteins are stored at 4 °C in 50 mM NaCl, 30 mM sodium cacodylate, pH 7.0, 10% glycerol, and ^{19}F -Tyr labeled proteins are stored in NMR buffer with 0.01% $\text{Na}_2\text{S}_2\text{O}_3$ at 4 °C.

Protein unfolding

Stabilities of RBDs were determined by chemical denaturation in GdnHCl, monitored by CD spectroscopy. All stability measurements were performed at 22 °C in 50 mM NaCl, 10 mM sodium cacodylate, pH 7.0, with varying concentrations of GdnHCl. Ultra-pure GdnHCl was used without further purification to make 8 M stock solutions containing NaCl and sodium cacodylate. The concentration of GdnHCl in the stock solution was verified by refractive index (Pace, 1986) prior to each experiment. Individual GdnHCl dilutions were made by mixing varying volumes of the stock solution containing the denaturant and a solution containing only NaCl and sodium cacodylate. A fixed volume of each RBD was added, between 30 μM and 50 μM protein, for each concentration of GdnHCl. At least

30 separate GdnHCl concentrations from 0 to 8 M were used for each experiment. GdnHCl-induced unfolding was measured with CD at 223 nm using a JASCO J-600 spectropolarimeter with a 1-mm circular cuvette. The ellipticity of individual samples was measured for 200 time points. Duplicate experiments for each RBD were averaged together. These averaged values of ellipticity were fit to the following equation using nonlinear least-squares analysis, based on the linear extrapolation method (Santoro & Bolen, 1988), using Kaleidegraph software.

$$y = \frac{(y_D)\exp[(mC - \Delta G_D^0)/RT] + (y_N)}{1 + \exp[(mC - \Delta G_D^0)/RT]}$$

$\Delta G_D = \Delta G_D^0 - mC$, where C is the concentration of GdnHCl, m is the slope of the transition; y is the observed molar ellipticity, and y_N and y_D are the native and denatured baseline linear functions, respectively (Pace, 1986).

Reversibility of unfolding was tested by measuring both molar ellipticity and RNA affinity of proteins used in stability studies. Proteins were refolded by adjusting the concentration of GdnHCl to 5.5 M at approximately 20 μM [protein], to ensure completely unfolded protein as the starting material. Proteins were then dialyzed against a reverse GdnHCl gradient from 5 M to 1.5 M GdnHCl in 50 mM NaCl, 10 mM sodium cacodylate, pH 7.0, at 4 °C, followed by dialysis against buffer containing 50 mM NaCl, 30 mM sodium cacodylate, pH 7.0, 10% glycerol. Spectra of refolded proteins have the same molar ellipticity as the original protein. Affinity of refolded proteins for wild-type RNA substrate is identical to purified protein, within error of the filter binding experiment.

Filter binding assays

Nitrocellulose filter binding was used to determine dissociation constants for the interaction between the wild-type RNA substrate and RBD1 constructs as described previously (Hall, 1994; Hall & Kranz, 1995). All binding assays were performed in 100 mM NaCl, 10 mM sodium cacodylate, 1 mM MgCl_2 , pH 7.0, at 22 °C, with a constant, trace concentration of [α - ^{32}P]-CTP, UTP-labeled RNA. Equilibrated samples were filtered over pre-soaked Schleicher & Schuell 0.2- μm BA-S 83 filters. Bound radio-labeled RNA was quantified from dried filters using a Betagen BetaScope (Model 603). The fractional saturation of RNA was determined by normalizing binding isotherms to total RNA concentration, then data were fit to a Langmuir isotherm by nonlinear least-squares in Kaleidegraph software. Retention efficiency ranged from 40 to 60%, typical for this RNA-protein interaction (Hall, 1994); data from multiple experiments were normalized and averaged for determination of dissociation constants. Reported dissociation constants represent an average of at least two experiments; data typically span a twofold range in K_D for individual experiments.

NMR spectroscopy

All NMR experiments used a Varian Unity-500 spectrometer with a Nalorac 5-mm dual coil HF500-5 probe or a Nalorac 5-mm triple resonance probe with z-gradients. All data from ^{19}F experiments were processed using Varian VNMR software. Proton chemical shifts were referenced to water at 4.8 ppm; fluorine chemical shifts were referenced to TFA at 0 ppm (in

$^1\text{H}_2\text{O}$ at 25 °C). 1D ^{19}F spectra were acquired with a spectral width of 8,000 Hz.

2D ^{19}F -detected HOESY experiments were performed on ^{19}F -Tyr RBD1 wild-type and variant proteins. The HOESY pulse sequence used in these experiments was similar to those employed by others (Rinaldi, 1983; Yu & Levy, 1983, 1984), except no decoupling or field gradient pulses were used (Cistola & Hall, 1995). The HOESY spectra were collected with spectral widths of 4,000 Hz (^{19}F) and 10,000 Hz (^1H). Protein spectra were acquired on 2 mM ^{19}F -Tyr RBD1, with a mixing time optimized to 300 ms and a 1.2-s interpulse delay. Spectra of the complex were acquired with 1 mM RNA: ^{19}F -Tyr RBD1, with a 175-ms mixing time and a 1.2-s relaxation delay.

For formation of the RNA:protein complex used in NMR experiments, ^{19}F -Tyr RBD1 at 200 μM in 100 mM NaCl, 10 mM sodium phosphate, 1 mM MgCl_2 , pH 6, was lyophilized and resuspended in $^2\text{H}_2\text{O}$ (lyophilization at higher starting protein concentrations can result in some protein remaining as precipitate). Separate experiments have shown lyophilization of RBD1 does not affect affinity of RBD1 for RNA (data not shown). The five-base pair stem RNA at 200 μM in NMR buffer was dried and resuspended in $^2\text{H}_2\text{O}$. The complex was formed by incremental addition of RNA to the protein to a final complex concentration of 0.1 mM (with 5% excess RNA), which was then concentrated in a 3000 MWCO Centricon (Amicon) at 4 °C to 1 mM for NMR experiments.

Dynamics measurements

^{15}N T_1 , T_2 , and NOE spectra were recorded at 25 °C using pulse sequences described by Farrow et al. (1994) using 1.5 mM ^{15}N uniformly labeled 1–102 amino acid protein. T_1 values were calculated from spectra recorded with seven different delay times: $T = 5.6, 56, 112, 224, 448, 728$, and 1,120 ms. T_2 values were determined from spectra recorded with delays $T = 16, 32, 48, 80, 112, 144$, and 176 ms. Duplicate spectra were recorded for $T = 5.6$ and 448 ms (T_1) and $T = 16$ and 112 ms (T_2) to estimate the noise levels. ^1H - ^{15}N steady-state NOE values are obtained by recording spectra with (NOE) and without (NO-NOE) the use of ^1H saturation applied before the start of the experiment. All of these spectra were recorded with 128×512 complex matrices with 64 scans per t_1 point. Spectral widths of 1,520 and 6,500 Hz were employed in F1 and F2, respectively. ^1H - ^{15}N HSQC spectra were also recorded with the same sample at 25 °C and 30 °C. All spectra were processed identically with Lorentzian–Gaussian weighting functions applied in both dimensions.

All spectra were processed with the hypercomplex method (States et al., 1982) using Varian software on a Sun Sparc Workstation. Analysis and peak picking were carried out using NMR Compass (Molecular Simulation Incorporated) on an SGI Indigo.

Backbone dynamic data analysis

Peak heights of 2D spectra were obtained using NMR Compass. T_1 and T_2 values were obtained from the fit of the intensity as a function of time to the following equation using Kaleidagraph software:

$$I(t) = I(0)\exp(t/T_{1,2})$$

where $I(t)$ and $I(0)$ is the intensity of the peak after delay of t and $t = 0$, respectively. Steady-state NOE values were determined from the ratios of the intensities with or without proton saturation.

Acknowledgments

We thank Tom Stump for excellent technical assistance, Dr. C. van Gelder (Nijmegen) for the reconstructed RBD1 coordinates, Prof. Linda Nicholson (Cornell) for NMR relaxation analysis software, and Dr. Gabriele Varani (M.R.C.) for providing the assignments of the N-terminal RBD. This work is supported by the NIH (GM46318), the Council for Tobacco Research, and the Lucille P. Markey Charitable Trust (#90-47) to K.B.H., and NIH GM16739 (J.L.).

References

- Bandziulis RJ, Swanson MS, Dreyfuss G. 1989. RNA-binding proteins as developmental regulators. *Genes & Dev* 3:431–437.
- Birney E, Kumar S, Krainer AR. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* 21:5803–5816.
- Blaber M, Zhang XJ, Lindstrom JD, Pepiot SD, Baase WA, Matthews BW. 1994. Determination of α -helix propensity within the context of a folded protein. *J Mol Biol* 235:600–624.
- Boelens WC, Scherly D, Beijer RP, Jansen EJR, Dathan NA, Mattaj JW, van Venrooij W. 1991. Analysis of in vitro binding of U1A protein mutants to U1 snRNA. *Nucleic Acids Res* 19:4611–4618.
- Burley SK, Petsko GA. 1986. Amino–aromatic interactions in proteins. *FEBS Lett* 203:139–143.
- Carne TJ, Huber RE, Davitt P, Edwards LA. 1986. Analysis of fluorine and iodine derivatives of tyrosine. *J Chromat* 367:393–403.
- Chakrabarty A, Kortemme T, Baldwin RL. 1994. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci* 3:843–852.
- Cistola DP, Hall KB. 1995. Probing internal water molecules in proteins using two-dimensional ^{19}F - ^1H NMR. *J Biomol NMR* 5:415–419.
- Dettman HD, Weiner JH, Sykes BD. 1985. A ^{19}F nuclear magnetic resonance study of the interaction of carbon dioxide with fluoro-amino acids. *Can J Biochem Cell Biol* 63:1120–1126.
- Dougherty DA. 1996. Cation- π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* 271:163–168.
- Dreyfuss G, Matunis MJ, Pinol-Roma S, Burd C. 1993. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* 62:289–321.
- Dreyfuss G, Swanson MS, Pinol-Roma S. 1988. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci* 13:86–91.
- Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. 1994. Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 33:5984–6003.
- Flocco MM, Mowbray SL. 1994. Planar stacking interactions of arginine and aromatic side-chains in proteins. *J Mol Biol* 235:709–717.
- Garrett DS, Lodi PJ, Shamoo Y, Williams K, Clore GM, Gronenborn A. 1994. Determination of the secondary structure and folding topology of an RNA binding domain of mammalian hnRNP A1 protein using three-dimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry* 33:2852–2858.
- Gerig JT. 1978. Fluorine NMR of proteins. In: Berlinger L, Reuben J, eds. *Biological magnetic resonance*. New York: Plenum Press. pp 139–203.
- Gorlach M, Wittekind M, Beckman RA, Dreyfuss G. 1992. Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *EMBO J* 11:3289–3295.
- Hall KB. 1994. Interactions of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry* 33:10076–10088.
- Hall KB, Kranz JK. 1995. Thermodynamics and mutations in RNA–protein interactions. *Methods Enzymol* 259:261–281.
- Hall KB, Stump WT. 1992. Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucleic Acids Res* 20:4283–4290.
- Higuchi R. 1990. Recombinant PCR. In: Innis MA, ed. *PCR protocols: A guide to methods and applications*. San Diego, California: Academic Press, Inc. pp 177–183.
- Hoffman DW, Query CC, Golden BL, White SW, Keene JD. 1991. RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc Natl Acad Sci USA* 88:2495–2499.
- Hull WE, Sykes BD. 1975. Fluorotyrosine alkaline phosphatase: internal mo-

- bility of individual tyrosines and the role of chemical shift anisotropy as a ^{19}F nuclear spin relaxation mechanism in proteins. *J Mol Biol* 98:121–153.
- Jorgensen ED, Durbin RK, Risman SS, McAllister K. 1991. Specific contacts between the bacteriophage T3, T7, and SP6 RNA polymerases and their promoters. *J Biol Chem* 266:645–651.
- Jessen TH, Oubridge C, Teo CH, Pritchard C, Nagai K. 1991. Identification of molecular contacts between the U1A small nuclear ribonucleoprotein and U1 RNA. *EMBO J* 10:3447–3456.
- Keene JD, Query CC. 1991. Nuclear RNA-binding proteins. *Prog Nucleic Acid Res* 41:179–202.
- Kenan DJ, Query CC, Keene JD. 1991. RNA recognition: Towards identifying determinants of specificity. *Trends Biochem Sci* 16:214–220.
- Lee AL, Kanaar R, Rio DC, Wemmer DE. 1994. Resonance assignments and solution structure of the second RNA-binding domain of Sex-lethal determined by multidimensional heteronuclear magnetic resonance. *Biochemistry* 33:13775–13786.
- Lu J, Hall KB. 1995. An RBD that does not bind RNA: NMR secondary structure determination and biochemical properties of the C-terminal RNA binding domain from the human U1A protein. *J Mol Biol* 247:739–752.
- Lutz-Freyermuth C, Query CC, Keene JD. 1990. Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc Natl Acad Sci USA* 87:6393–6397.
- Mattaj JW. 1989. A binding consensus: RNA-protein interactions in splicing, snRNPs, and sex. *Cell* 57:1–3.
- Merrill BM, Stone KL, Cobianchi F, Wilson SH, Williams KR. 1988. Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein. *J Biol Chem* 263:3307–3313.
- Mitchell JBO, Nandi CL, McDonald IK, Thorton JM, Price SL. 1994. Amino/aromatic interactions in proteins: Is the evidence stacked against hydrogen bonding? *J Mol Biol* 239:315–331.
- Nagai K, Oubridge C, Jessen TH, Li J, Evans PR. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* 348:515–520.
- O'Neil KT, DeGrado WF. 1990. A thermodynamic scale of the helix-forming tendencies of the commonly occurring amino acids. *Science* 250:646–651.
- Orengo CA, Thorton JM. 1993. Alpha plus beta folds revisited: Some favoured motifs. *Structure* 1:105–120.
- Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 372:432–438.
- Pace CN. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131:266–280.
- Query CC, Bentley R, Keene J. 1989. A common RNA recognition motif identified within a defined U1 RNA-binding domain of the 70K U1 snRNP protein. *Cell* 57:89–101.
- Rinaldi PL. 1983. Heteronuclear 2D-NQE spectroscopy. *J Am Chem Soc* 105:5167–5168.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. Chapter 14.
- Santoro M, Bolen DW. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α -chymotrypsin using different denaturants. *Biochemistry* 27:8063–8068.
- Scherly D, Boelens W, Dathan NA, van Venrooij WJ, Mattaj IW. 1990. Major determinants of the specificity and interaction between small nuclear ribonucleoproteins U1A and U2B⁺ and their cognate RNAs. *Nature* 345:502–506.
- Scherly D, Boelens W, van Venrooij WJ, Dathan NA, Hamm J, Mattaj IW. 1989. Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J* 8:4163–4170.
- Sillikens PTG, Habets WJ, Beijer RP, van Venrooij WJ. 1987. cDNA cloning of the human U1 snRNA-associated A protein: Extensive homology between U1 and U2 snRNP-associated proteins. *EMBO J* 6:3841–3848.
- States DJ, Haberkorn RA, Ruben DJ. 1982. A two-dimensional nuclear Overhauser experiments with pure absorption phase in four quadrants. *J Magn Reson* 48:286–292.
- Stump WT, Hall KB. 1993. SP6 RNA polymerase efficiently synthesizes RNA from short double-stranded DNA templates. *Nucleic Acids Res* 21:5480–5484.
- Stump WT, Hall KB. 1995. Crosslinking of an iodo-uridine-RNA hairpin to a single site of the human U1A N-terminal RNA binding domain. *RNA* 1:55–63.
- van Gelder CWG, Gunderson SI, Jansen EJR, Boelens WC, Polycarpou-Schwartz M, Mattaj IW, van Venrooij WJ. 1993. A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A is required for regulation of polyadenylation. *EMBO J* 12:5191–5200.
- van Gelder CWG, Leusen FJJ, Leunissen JAM, Noordik JH. 1994. A molecular dynamics approach for the generation of complete protein structures from limited coordinate data. *Proteins Struct Funct Genet* 18:174–185.
- Wittekind M, Gorchach M, Friedrichs M, Dreyfuss G, Mueller L. 1992. ^1H , ^{13}C , and ^{15}N NMR assignments and global folding pattern of the RNA-binding domain of the human hnRNP C proteins. *Biochemistry* 31:6254–6265.
- Wojcik J, Altmann KH, Scheraga HA. 1990. Helix-coil stability constants for the naturally occurring amino acids in water. XXIV. *Biopolymers* 30:121–134.
- Yu C, Levy GC. 1983. Solvent and intramolecular proton dipolar relaxation of the three phosphates of ATP: A heteronuclear 2D NOE study. *J Am Chem Soc* 105:6994–6996.
- Yu C, Levy GC. 1984. Two-dimensional heteronuclear NOE (HOESY) experiments: Investigation of dipolar interactions between heteronuclei and nearby protons. *J Am Chem Soc* 106:6533–6537.